

Transistor Basics

And Their Role Today

By **Stuart Ball**

In this day and age of highly integrated chips, what is the relevance of the lone, discrete transistor? It's true that most embedded system design needs can be met by chip-level solutions. But electronic component vendors do still make and sell individual transistors because there's still a market for them. In this article, Stuart reviews some important basics about transistors and how you can use them in your embedded system design.

What good is a transistor? Sure, integrated circuits (ICs) are full of transistors, thousands of them. Before the IC and microprocessor revolutions, there was a transistor revolution—where televisions, radios and computers were built using the new solid-state devices. The transistor was the father of the IC. But isn't a single transistor obsolete as a circuit element today? What use does a lowly transistor have in a world where the current Intel microprocessors have over a billion transistors each?

It's true that nearly all the things we used to do with transistors can be done cheaper, better and more efficiently with an IC, and we can do things with ICs that are not possible with discrete transistors. It would not be possible to build a modern microprocessor

with discrete transistors—the lead lengths alone would make the speeds impossible. But the reverse is also true. A discrete transistor can be a simple way to solve some problems. Transistors, for example, typically have much higher operating voltage and power limits in simple circuits than those of comparable ICs. Electronics manufacturers and distributors still make and sell individual transistors because the parts still have some uses. In this article, I want to go over some very basic things about transistors, how they are used and how you can include them in your applications.

OVERVIEW

A BJT (bipolar junction transistor) was the first commonly available transistor, and it fueled the transition away from vacuum tubes. BJTs come in two varieties, NPN and PNP. Both are (usually) silicon devices. The silicon is modified (doped) with impurities to produce N-type or P-type material. An NPN transistor has a P-type layer sandwiched between two N-type layers, and a PNP is the reverse.

Figure 1 shows an NPN BJT schematic symbol, a simple diagram of the structure, and a diode model. The N-P-N structure is just representative. In an actual transistor, the collector region is normally larger than the emitter region, and none of them is square as shown in the diagram. The diode representation of the transistor indicates how current flows, not how the actual part is constructed. You can't build a transistor out of two diodes, but using two diodes helps to explain how the transistor biasing works.

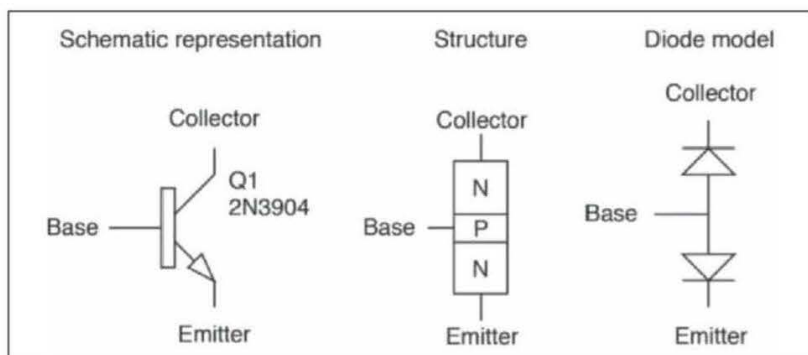


FIGURE 1
Schematic symbol, physical representation and diode model of NPN transistor

Operation of an NPN transistor is conceptually easy to understand. Referring to the diode model, if you connect the collector to a positive voltage—say 5 V—and the emitter to ground, you end up with two diodes back-to-back with their anodes connected together. The junction of the two anodes represents the base of a transistor. If you apply a positive voltage greater than 0.7 V to the base, the emitter diode will be forward-biased and current will flow from the base, through the emitter and to ground. The collector diode will be reverse-biased, and no current will flow through that diode.

REAL TRANSISTOR OPERATION

Now discard the diode model and look at a real transistor. If the collector is connected to +5 V and the emitter to ground, and the voltage on the base is high enough (0.7 V) to forward-bias the base-emitter junction, current will flow from the base to the emitter *and* from the collector to the emitter. If the base-emitter voltage is below 0.7 V, the transistor is in “cutoff” and no current flows through the emitter or through the collector. That’s it. That’s how a BJT works.

The collector-emitter current flow is inherent in the construction of the transistor. It’s why the actual transistor differs from the diode model, and it’s why you can’t build a transistor from two diodes. If the collector is at +5 V and the emitter is at ground, bringing the base to about 0.7 V will cause current to flow from the 5 V supply—through the collector—to the emitter and to ground. If the emitter is at +2 V, then you must bring the base to about 2.7 V to get current to flow from the collector to the emitter.

The magic in a transistor is determining how to get the amount of current you want flowing through the collector. If you just connect the transistor as I’ve described, with nothing to limit the current, your transistor will quickly become a smoking, melted bit of plastic.

Generally, if the transistor is operated within its current, power and voltage ratings, the current in the emitter will be the current flowing into the base plus the current flowing from the collector to the emitter. A very small base current controls a much larger collector current, so the collector current is approximately equal to the emitter current. When no current is flowing in the collector, the transistor is in “cutoff” as mentioned earlier. If there is enough current flowing that the collector-emitter voltage is as low as it can go (generally around 0.3 V for a small-signal transistor), the transistor is considered “saturated”. In this state, changes to base current no longer affect collector current.

PUTTING IT TO USE

How might we use this transistor? **Figure 2** shows a simple circuit. In this circuit, we connect

the collector to +5 V, the emitter to ground through a 220 Ω resistor and the base to a fixed value of 1 V. The forward voltage of the 2N3904 is 0.65 V to 0.85 V at 10 mA collector current. Conventionally, 0.7 V is used for calculations. So, the voltage at the emitter (VE) will be 1 V - 0.7 V, or 0.3 V. Here’s where the magic happens: The voltage at the emitter is fixed, so the current through the 220 Ω resistor is $0.3\text{V}/220\Omega$, or 1.36 mA. The collector current is the same. Therefore, by controlling the base voltage, we control the emitter current and thereby the collector current.

Figure 3 shows how we can make an amplifier with this circuit. This circuit is identical to the circuit in Figure 2, except that now we’ve added a 1.5 k Ω resistor, R2, between the collector and the 5 V supply. Since the current in the emitter is fixed at 1.36 mA, the current in the collector is also 1.36 mA. This current flows through R2, producing a voltage across R2 of $1.36\text{ mA} \times 1.5\text{ k}\Omega$, or 2.04 V. So, the voltage at the collector, VC, is the 5 V supply minus the voltage across R2, or 2.95 V.

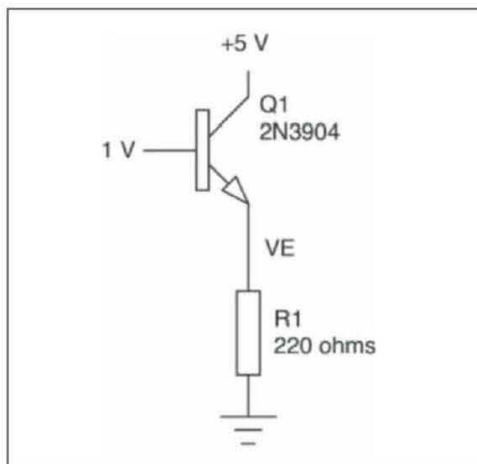


FIGURE 2

A simple circuit shows the base-emitter voltage and current relationship

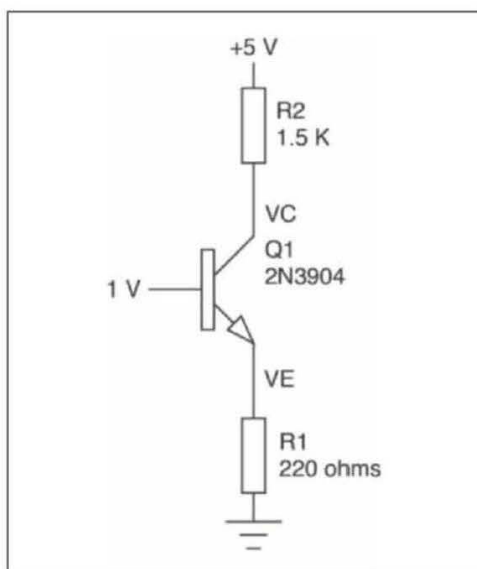


FIGURE 3

The transistor connected as an amplifier by adding a resistor in the collector

Now, what happens if the voltage at the base is raised to 1.1 V? When that happens, the voltage at the emitter is now 0.4 V ($1.1 \text{ V} - 0.7 \text{ V}$), making the emitter current 1.8 mA. The collector current is also 1.8 mA, so the voltage across R2 is now $1.8 \text{ mA} \times 1.5 \text{ k}\Omega$, or 2.7 V. VC is now $5 \text{ V} - 2.73 \text{ V}$, or 2.27 V. So, a 0.1 V change in the base voltage caused the collector voltage to drop from 2.95 V to 2.27 V, a change of -0.68 V. The collector voltage dropped by 6.8 x 0.1 V (the input voltage change).

Here's the interesting thing: The collector voltage change is equal to the negative of the input voltage change times the ratio of the collector resistor R2 to the emitter resistor R1, or $1.5 \text{ k}\Omega / 220 = 6.8$. If you work through the math, this makes sense, because the collector current is the same as the emitter current. But since the collector resistor R2 is 6.8x the emitter resistor, any current change in the emitter resistor will result in a voltage change 6.8x as large at the collector.

If you did the same calculation after lowering the base voltage from 1 V to 0.9 V, you would see the collector voltage rise by 0.68 V. This circuit is an inverting amplifier

with a gain of -6.8. A positive voltage change at the input produces a negative voltage change at the output and vice-versa.

This circuit has some limitations. If you put 1.32 V at the base, you will find that the emitter is at 0.62 V, and the collector voltage works out to be nearly the emitter voltage. The transistor can't drive the collector to the emitter voltage, so it's saturated. The limitation of this specific circuit, therefore, is a maximum input voltage of about 1.3 V. At the other end, anything less than 0.7 V causes the transistor to go into cutoff. So, the useful input voltage range of this circuit is 0.7 V to about 1.3 V. Still, that would be adequate for boosting a low-level audio signal to something that can be further amplified.

Speaking of audio, how would you connect audio signals into the circuit? Audio signals typically swing between negative and positive voltages. If you put that into the base, the transistor will be in cutoff most of the time—all the time if the positive signal peaks never reach 0.7 V.

This brings us to biasing. **Figure 4** is a modification of Figure 3 with some biasing resistors added to the base. Resistors R3 and R4 make a voltage divider that brings the base to about 1 V. This is halfway between the 0.7 V and 1.3 V lower and upper limits of the circuit. Now say that we apply a signal to the input that swings between -0.1 V and +0.1 V. Because of the DC blocking capacitor C1, this will become 0.9 V to 1.1 V at the base, and will be amplified by -6.8 in the circuit.

There are other ways to bias a transistor base. A voltage reference diode, as shown in **Figure 5**, fixes the base at a known voltage. In this circuit, the emitter voltage, VE, will be about 1.3 V, so the emitter and collector current will be 5.9 mA. The point is not to show all the possible ways to bias a transistor, just that there are other ways to do it.

TRANSISTOR LIMITATIONS

As with all things in the physical world, transistors have some limitations. We already looked at one—the values of the base and emitter resistors in the amplifier circuit have to be chosen so that the transistor doesn't go into cutoff or saturation with whatever input signal you are trying to amplify.

Transistors have other characteristics. For example, the 2N3904 used in these examples has a maximum collector-emitter voltage of 40 V. Any more than that, and the transistor fries. The base-emitter reverse voltage—where the base is taken negative with respect to the emitter—has a maximum value of 6 V. Beyond that, the emitter-base junction breaks down.

The collector can handle a maximum continuous current of 200 mA. The device

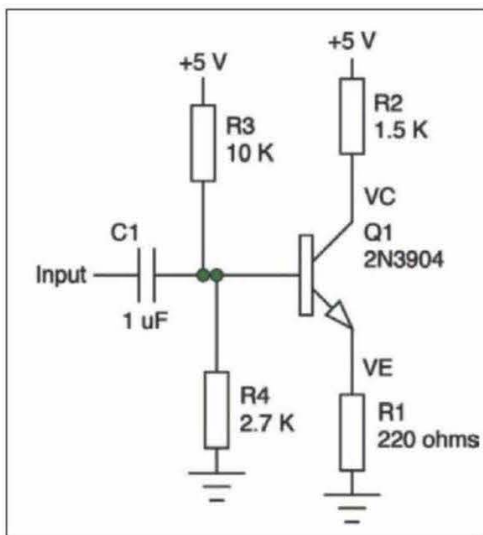


FIGURE 4

Biasing resistors allow the transistor to operate with AC-coupled inputs such as audio signals.

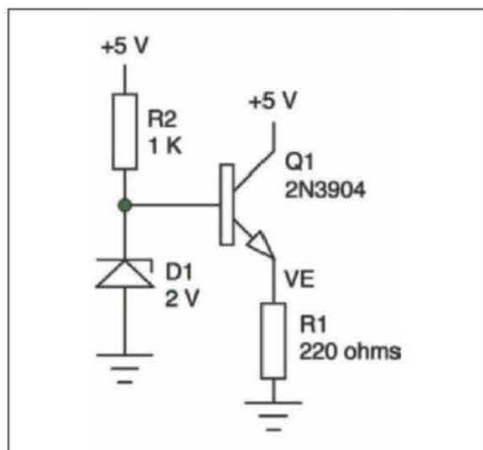


FIGURE 5

A Zener or reference diode can be used to create a fixed bias.

has a maximum power dissipation of about 600 mW. So even though the collector-emitter can withstand 40 V and the collector current can be as high as 200 mA, if you try to put 200 mA through it at 40 V, it will fail. 40 V at 200 mA is 8 W, well beyond the power-handling capability of the device.

The point of all this is that, like any semiconductor device, your design has to stay within all the maximum ratings: Power, collector-emitter voltage, collector current, emitter-base reverse breakdown voltage and so on.

One of the key characteristics of the transistor is the current gain. This number describes how much the emitter current changes for a given change in the base current. The current gain varies with the amount of current flowing in the collector. For the 2N3904, the minimum current gain at 0.1 mA collector current is 40. At 10 mA, the minimum gain is 50. The maximum gain per the datasheet is 300. Just before writing this paragraph, I measured a handful of 2N3904s. All of them had gain exceeding 300.

The practical implication of the gain is to affect how the emitter interacts with the base. If the transistor in the amplifier circuit in Figure 3 had a gain of only 10, the 220 Ω resistor in the emitter would look like approximately 2 k Ω at the base, which would affect biasing and the load presented to the driving circuit. In that case, you would want the biasing resistors to be a low enough value that the loading effect of the emitter resistor would change the bias voltage by less than 10% or so. But if you have to use lower value resistors in your biasing circuit, this in turn presents more load to whatever is driving it. In the case of the amplifier, it reduces the overall end-to-end gain.

Fortunately, for most small-signal applications, it isn't too hard to find a transistor with a sufficiently high minimum gain to make this a minor problem. Where you get into difficulty is when you need a very low value of emitter resistance. Even at a gain of 300, an emitter resistor of about 10 Ω could have a significant loading effect on the base that must be considered in your calculations. Because the transistor has finite gain, you can't use very large resistors—such as something in the megaohm range—to bias the base. If you do, the emitter will pull down the voltage.

One common addition to an audio amplifier is to bypass the emitter resistor with an electrolytic capacitor. The capacitor has a very high impedance (nearly infinite) at DC, but the impedance decreases as frequency increases. This allows the DC biasing to work, but it raises the gain for audio signals by making the emitter impedance (the resistance in parallel with the impedance of the capacitor) a very low value at audio frequencies. This makes the ratio of the collector resistance to emitter resistance much higher at audio than at DC, which raises the gain. (Remember: The gain is the collector resistor divided by the emitter impedance.) However, this also has the effect of significantly lowering the input impedance of the circuit at those audio frequencies. Other transistor characteristics that affect use in RF circuits, such as high-speed switching circuits, are beyond the scope of this article, and won't be discussed here.

APPLICATIONS

You can build amplifiers with transistors, and a lot of people do. But it's also easy to build an amplifier with an op amp or other IC and I want to focus here on applications where the unique characteristics of a transistor are useful.

How might you make practical use of a transistor, given what we've done so far? In **Figure 6**, I have modified Figure 5 by

FPGA Boards from HUMAN DATA JAPAN

SAVING COST=TIME with readily available FPGA boards

- Basic and simple features, single power supply operation
- Free download technical documents before purchasing

INTEL ACM-033

Intel Cyclone 10 LP F484 FPGA board



Cyclone10LP SDRAM SPI-Flash

SIZE : 3.386" x 2.126" (86 x 54 mm)

ACM-033 is an FPGA board with Intel high-performance FPGA Cyclone 10 LP. It's compact and very simple. 3.3V single power supply operation.

XILINX XCM-025

Xilinx Spartan-7 FGGA484 FPGA board



Spartan-7 DDR3

SIZE : 3.386" x 2.126" (86 x 54 mm)

XCM-025 is an FPGA board with Xilinx high-performance FPGA Spartan-7. It's compact and very simple. 3.3V single power supply operation.

See all our products, A/D D/A conversion board, boards with USB chip from FTDI and accessories at:

www2.hdl.co.jp/CC19B



making the reference voltage 2.5 V, making R1 120 Ω and adding an LED in the collector circuit. Because the voltage at the base is fixed at 2.5 V by the reference diode, the emitter voltage is 1.8 V and the emitter current is 15 mA. This is true as long as the V+ supply voltage is high enough to keep the reference diode and LED turned on. So, the LED will have 15 mA current whether the supply voltage is 5 V or 20 V.

Obviously, there are upper limits to this, and at some point, the voltage or power dissipation limit of the 2N3904 will be exceeded and it will go up in a cloud of smoke. I've shown the bias circuit powered from 5 V. If you also powered it from the variable V+, you also would need to take the limitations of R3 and D1 into account. But if you wanted a constant current through an LED regardless of supply voltage (within reasonable limits), this circuit will do it. You might do this if you wanted an LED to have constant intensity regardless of the voltage applied, or just to keep higher voltages from exceeding the maximum LED current.

Figure 7 shows a 2N3904 used for logic-level translation between two different circuits operating at different voltages. You might use

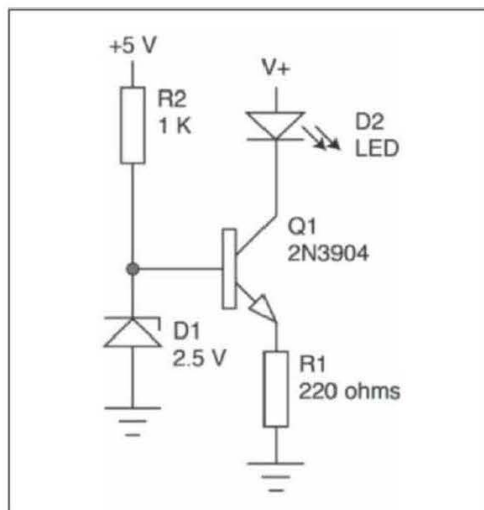


FIGURE 6

A 2N3904 connected as a constant-current LED driver

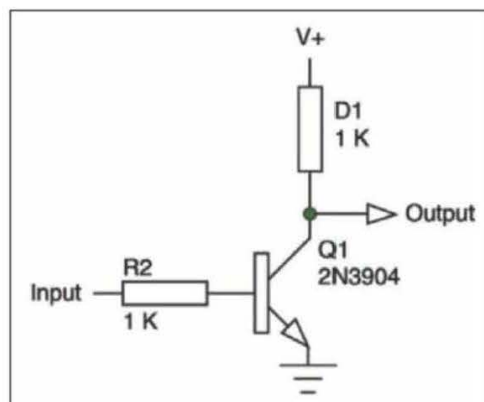


FIGURE 7

A 2N3904 used as a logic-level translator

this to translate between a 3.3 V output of a microcontroller (MCU) to the input of a circuit that needs 5 V. V+ in the schematic would be connected to the supply voltage of the target system. Whatever is driving the input must have enough output current capability to drive the 2.2 k Ω resistor. This circuit inverts the signal—a high input produces a low output. In this circuit, the transistor is always in either cutoff or saturation.

There are plenty of ICs that can do this, such as open-collector buffers, so why use a transistor? The transistor can handle higher voltages than most logic-level translator circuits. A transistor could translate between a 3.3 V circuit and a 12 V circuit, for example.

Many voltage-translator circuits require that you know the supply voltage, and therefore the drive voltage, of the input. But I had a situation once where the input could come from different sources, ranging from under 2.5 V to 5 V. The transistor solution works for all logic voltages, because the transistor will turn on with any drive voltage above 0.7 V. It could even be used to translate between a 12 V or 24 V input to a 3.3 V or 5 V output, as long as the input resistor R2 is large enough to prevent excessive current.

The final NPN application is shown in **Figure 8**. In **Figure 8a**, a 2N3904 is driving a relay. The diode D1 protects the transistor against overvoltage. When the relay is turned off by switching the transistor off, a “flyback” voltage is created as the energy in the relay coil is dissipated. This voltage can reach levels sufficient to destroy the transistor due to excessive collector-emitter voltage—remember the transistor characteristics section. Diode D1 limits the voltage to 0.7 V above V+ to protect the transistor. But this has the side effect of slowing down relay opening.

Figure 8b shows the same circuit, but with a 12 V Zener, D2, in series with D1. This allows the flyback voltage to reach 12.7 V above V+, which allows the coil energy to be dissipated much more quickly, speeding up relay operation. But with a 12 V relay, the collector voltage will exceed 24 V during the flyback period. This circuit takes advantage of the high collector-emitter breakdown voltage to improve the speed. There are some relay drivers that can do this, but they offer little advantage over a transistor. Note however that base resistor R1 must be sized to allow enough current for the transistor to operate the relay. A large, high-current relay may require a pre-driver and a power transistor. At that point, an IC might be a better solution.

PNP TRANSISTORS

I've focused on NPN transistors so far. Functionally, the PNP is the reverse of the

NPN. The collector voltage of the PNP (when normally biased) is less than the emitter, and the base is lower than the emitter by 0.7 V to turn the transistor on. It isn't necessary to use negative voltages. As with the NPN, the voltage with respect to the emitter is what matters. A PNP transistor can be paired with an NPN in simple audio amplifiers to make a headphone or speaker amplifier. The PNP complement to the 2N3904 is the 2N3906.

Figure 9 shows how a 2N3906 might be used to create a negative bias voltage in a system with only a positive supply. You might need a negative bias to offset an input signal, or to power an op-amp that needs a negative supply for some reason.

The input is driven by a square wave input that might come from the timer output of a MCU or a two-transistor multivibrator (Google it). I picked values arbitrarily for the components in this example. You would want to use component values appropriate for the input frequency, output current and voltage, and other requirements of your application. Note that the input signal must swing close to the positive supply rail (5 V in the circuit shown) to fully turn off Q1—otherwise the transistor will never turn off, and it will get hot. If you were driving the circuit with a logic-level output, you might need a pull-up resistor to be sure the input swings all the way to the positive rail. You could also use this circuit in a 3.3 V system.

I include this example to show how a PNP transistor can be used. This isn't to say there aren't ICs that can do this. For example, the TPS6735 DC/DC converter made by Texas Instruments can produce a -5 V output at 200 mA, although it won't operate at 3.3 V.

MOSFET TRANSISTORS

I've looked at BJTs so far, but there is another class of transistors called MOSFETs (metal-oxide semiconductor field effect transistors). Where a BJT has a base, emitter

ABOUT THE AUTHOR

Stuart Ball is a registered professional engineer with a BSEE and an MBA. He has more than 30 years of experience in electronics design. He is currently a principal engineer at Seagate Technologies.

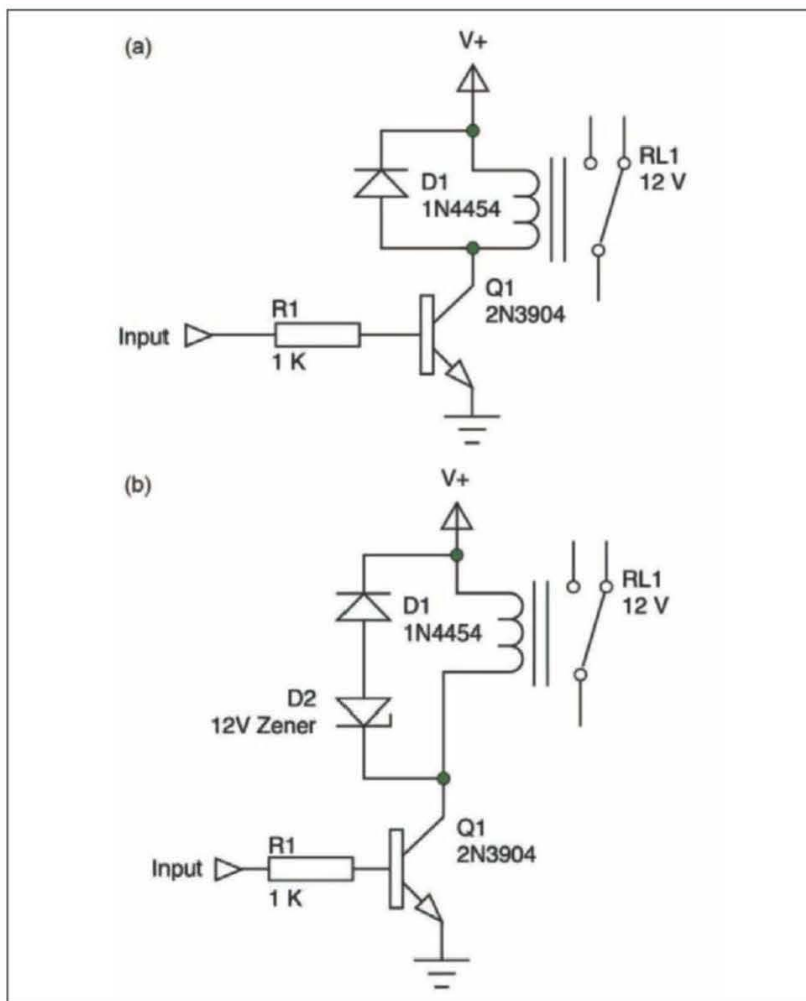


FIGURE 8

Driving a relay with a 2N3904. A basic diode clamp (a) and a higher voltage Zener clamp (b) for faster operation.

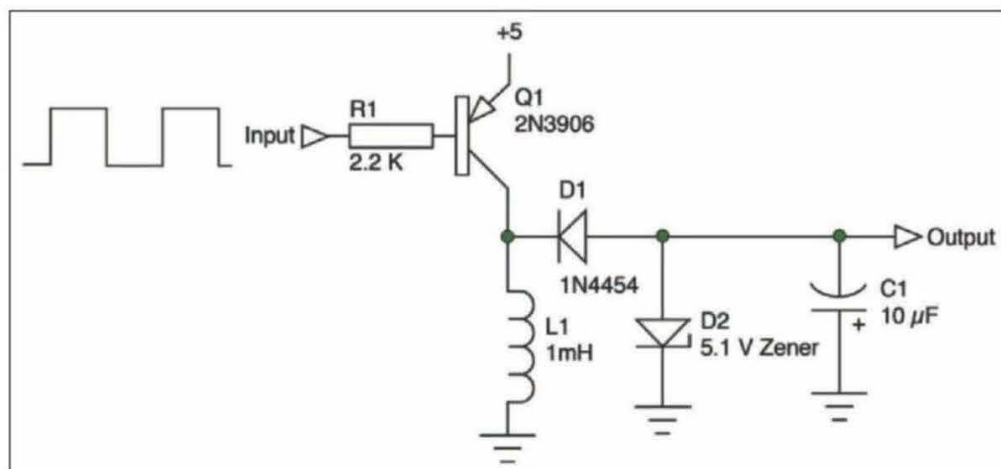
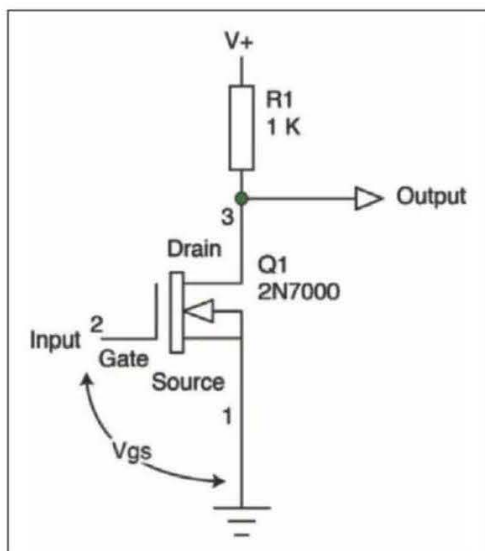


FIGURE 9

Negative voltage generator using PNP 2N3906

FIGURE 10

2N7000 MOSFET as inverting logic-level converter



and collector, the equivalent MOSFET pins are the gate, source and drain. MOSFET operation is similar to the BJT, but there are some important differences.

The MOSFET was sometimes previously referred to as the IGFET (insulated-gate field effect transistor). I haven't seen that term used for many years, but it is descriptive. The gate of the MOSFET is electrically insulated from the rest of the part, and the current from the drain to the source is controlled by the electrical field created by applying a voltage to the gate. The insulated gate means that the MOSFET has a very high impedance input, so no current has to flow into the gate to control the drain-source current. In fact, if current is flowing into the gate, it probably means that some limit has been exceeded and the transistor has failed.

The BJT can be thought of as a current-controlled current device, where a small change in base current causes a large change in collector current. A MOSFET is a voltage-controlled current device, where a change in the gate voltage causes a large change in the drain current. **Figure 10** shows a 2N7000 MOSFET connected as a logic-level translator, similar to the way the BJT was wired in **Figure 7**. It will work the same way as the 2N3904 circuit, with the following differences:

1. The high impedance means no series resistor is needed in the gate to limit current. This also means that the transistor input won't load down whatever output is driving it.
2. The BJT needs 0.7 V and a little current to turn the transistor on. The MOSFET needs the gate to be positive with respect to the source. In the case of the 2N7000, the turn-on voltage, V_{gs} , can range from 0.8 V to 3 V. This means that using a 2N7000 to translate between a 2.5 V or 3.3 V input to a higher voltage output might be problematic, and the transistor might not turn on. However, going the other way, from a 5 V or higher system input to 3.3 V or 2.5 V output, will work the same as it does with the bipolar circuit.
3. A saturated MOSFET doesn't have a saturation voltage—it has a resistance between the source and drain. For the 2N7000, this can be up to about $6\ \Omega$ when $V+$ is 5 V for the On Semiconductor version of the part. For most applications, this value is small enough that it makes no difference, but it is something to be aware of, especially when switching significant current.

The 2N7000 is normally used as a switch. You can bias it as an amplifier, but the varying V_{gs} threshold value makes that a bit more complicated than for a BJT. Like the PNP complement to the NPN transistor, N-channel MOSFETs have a complement, which is the P-channel MOSFET. The BS250 from Vishay is an approximate P-channel equivalent to the 2N7000. You could use such a transistor instead of a PNP to implement the negative voltage generator mentioned earlier, although, of course, you have to be sure the driving voltage exceeds the gate threshold voltage.

OTHER TRANSISTORS

I've focused on small-signal transistors to demonstrate the basic principles. In both bipolar and MOSFET transistors there are devices designed to handle high currents and high voltages, parts designed specifically for RF applications, and other variants. But the basic principles are the same.

I hope my explanation of how transistors work has helped you understand them better, and that the examples are enough to let you experiment with transistors in your applications. Sometimes transistors are useful, even though they've been around a long time. And even in circuits you could build with ICs, transistors are interesting devices for tinkering, because you can get down to the basic component level. 

For detailed article references and additional resources go to:
www.circuitcellar.com/article-materials

RESOURCES

On Semiconductor | www.onsemi.com

Texas Instruments | www.ti.com

Vishay | www.vishay.com